1st Pan-American Congress on Computational Mechanics - PANACM 2015
XI Argentine Congress on Computational Mechanics - MECOM 2015
S. Idelsohn, V. Sonzogni, A. Coutinho, M. Cruchaga, A. Lew & M. Cerrolaza (Eds)

A METHOD FOR ATOMS SELECTION APPLIED TO SCREENING FOR SLEEP DISORDERS

Román E. Rolon[1], Leandro E. Di Persia[1], Hugo L. Rufiner[1,2] and Rubén D. Spies[3]

[1]Instituto de Investigación en Señales, Sistemas e Inteligencia Computacional (sinc(i),UNL-CONICET), Fac. de Ingeniería y Cs. Hídricas, Univ. Nacional del Litoral, Santa Fe, Argentina.
Ciudad Universitaria CC 217, Ruta Nac. 168, km. 472.4 (3000), Santa Fe Argentina
E-mails: rrolon@santafe-conicet.gov.ar, ldipersia@fich.unl.edu.ar and lrufiner@gmail.com

[2]Laboratorio de Cibernética, Fac. de Ingeniería, Univ. Nacional de Entre Ríos, Argentina.

[3]Instituto de Matemática Aplicada del Litoral (IMAL), Santa Fe, Argentina.
Colectora Ruta Nac. 168, Km. 0 (3000), Paraje El Pozo, Santa Fe, Argentina
E-mail: rspies@santafe-conicet.gov.ar

Key words: Sparse Representations, Apnea-hypopnea, Inverse problems

Abstract. The Obstructive Sleep Apnea-Hypopnea Syndrome (OSAHS) is a sleep disorder which consists in repetitive events of partial or total airflow decrease during sleep. This pathology has a 4% prevalence in the population around the world and, without appropriate treatment, it increases with age. Actually the gold standard for detecting OSAHS is a polysomnography in a sleep laboratory, which consists in the simultaneous measurement of different physiological signals. In the last years several research studies have shown that the pulse oximetry is a very attractive option of screening for OSAHS, since changes in the dynamics of oxygen in the blood stream (SaO_2) can be related with respiratory problems.

In the last fifteen years, many different signal processing techniques were used for building appropriate representations of a certain types of signals. One of these techniques is known as "sparse representation". The idea behind the method is to represent the involved signal using only a few coefficients in a certain dictionary, previously constructed. In this work the SaO_2 signal is used in order to predict the occurrence of Apnea-Hypopnea (AH) events. First a dictionary is learned by using a statistical method (NOCICA), then a greedy pursuit algorithm is used in order to obtain the activation coefficients. A subset of the most discriminative coefficients is then selected and used as input of a pattern recognition neural network in order to classify AH events. The problem for finding the optimal dictionary and activation coefficients gives rise to an inverse problem with sparse

constraints. A multilayer perceptron with different number of inputs and neurons in its hidden layer is then tested and the optimal configuration is derived.

1 INTRODUCTION

The American Academy of Sleep Medicine (AASM) distinguishes more than 80 different sleep disorders [1]. One of those pathologies is the Obstructive Sleep Apnea-Hypopnea Syndrome (HS). This pathology is characterized by repetitive episodes of airway narrowing or collapse during sleep. Actually, this pathology affects between 2% to 4% of the population around the world and this prevalence increases with age [2] [3] [4].

The current diagnostic tool for detecting OSAHS is an overnight polysomnography (PSG) in a sleep laboratory whose accessibility in Argentina like in others countries is very limited. Pulse oximetry has become a very attractive option for detecting OSAHS, because its easy accessibility, cost effective and noninvasive and since a decrease in blood oxygen saturation (SaO_2) indicates respiratory problems. There is sufficient evidence that if this pathology is not tratted properly, the OSAHS is directly related to arterial hypertension, cardiovascular and cerebral vascular diseases, and also high risk of traffic accidents [5], [6].

To grade the OSAHS severity, an index called Apnea Hypopnea Index (AHI) is defined. AHI index is obtained by counting the total number of Apnea-Hypopnea (AH) events per hour while sleeping. A patient with an AHI less than 5 is considered *normal*, between 5 and 15 *mild*, between 15 and 30 *moderate* and more than 30 *severe* [3].

In the last fifteen years, many different approaches to traditional signal processing problems were taken. Some of these new formulations gave rise to techniques based on non-linear systems and higher-order statistics, including Independent Component Analysis (ICA) [7] and methods to obtain an Sparse Representation (SR) [8] [9] of a signal. They provide new ways of phrasing the problems of signal modeling and representation. One underlying idea is that of representing the involved signals using only a few significant characteristics, e.g. as an SR with just a few basic waveforms.

In a previous work a Most Discriminative Atom Selection (MDAS) [10] method was used in order to detect AH events by using only the SaO_2 signal. Now MDAS method is used for screening patients clinically suspected to having a moderate or severe OSAHS.

In this work we start by comparing the performances of an overcomplete assembled dictionary (OAD), trained using class information and a complete dictionary (CD), trained without class information. Those dictionaries were used as generators of an SR of the SaO_2 signal, preserving as much as possible the morphology of the signal. After that, the most discriminative activation coefficients, in certain sense, are selected and used to train a Multilayer Perceptron Neural Network (MLP) in order to detect the respiratory events.

The next section describes methods used to learn a dictionary and to inference the coefficients of an SR for a given signal. Preliminary results for diagnosing moderate or severe OSAHS are presented.

2 METHODS

2.1 SPARSE REPRESENTATION OF SIGNALS

By a dictionary we shall mean a matrix $\Phi \in \mathbb{R}^{N\times M}$ (with $M \geq N$) whose columns ϕ_j are called atoms. A representation of a signal $\mathbf{s} \in \mathbb{R}^N$ in terms of a fixed dictionary Φ can be stated as follow:

$$\mathbf{s} = \sum_{i=1}^{M} \phi_j a_j = \Phi\mathbf{a}, \tag{1}$$

where $\mathbf{a} = (a_j) \in \mathbb{R}^M$. Although inappropriately, the term “basis” instead of “dictionary” is sometimes used. Since the atoms are not required to be linearly independent and quite often more atoms that the space dimension are used, the latter situation is usually preferred.

When there are more columns in the dictionary than the size of $\mathbf{s}$, i.e. when $M > N$ (referred to as an overcomplete dictionary), or when the columns do not form a basis, then there may be non-unique representations of a given signal. In this situation a suitable criterion is required to select only one of those representations. In this context, sparsity often refers to the criterion of choosing a representation with few non-zero coefficients. In particular the SR problem of $\mathbf{s}$ can be stated as follows:

$$\mathbf{a}_{SR} = \underset{\mathbf{a}\in\mathbb{R}^M}{\operatorname{argmin}} ||\mathbf{a}||_0 \text{ subject to } \Phi\mathbf{a} = \mathbf{s}, \tag{2}$$

where $||\cdot||_0$ denotes the zero-norm.

It is important to note that although the mapping $\mathbf{s} \to \mathbf{a}$ in the representation (1) is obviously linear, if the dictionary Φ is overcomplete, the mapping $\mathbf{s} \to \mathbf{a}_{SR}$ is not necessary linear. Briefly said, under the sparsity condition (2) the mapping signal-to-coefficients may not be linear.

Let us consider now the problem of finding an SR of a given family of signals with respect to a fixed dictionary Φ, where learning such a dictionary is part of the problem. Clearly one could learn the dictionary using all the signals in the given family. Although this choice of Φ will result in optimal sparsity, most likely it will be highly undesired, mainly because of its size and redundancy. It becomes then necessary to find a dictionary that be optimal, in a certain sense, for a given family of signals.

2.2 FEATURE EXTRACTION

In a previous work [10] a method called Most Discriminative Atoms Selection (MDAS) is developed in order to improve the performance of a MLP. First MDAS selects the columns of a given dictionary that are mostly involved (activated) in the SaO_2 signal recovery. In this way a j^{th} column of a dictionary has an activation frequency ν_{ci}^{j} given the *class* i, where ν_{ci}^{j} make reference to the number of times that the j^{th} column is used for *class* i signal recovery.

The candidates to be considered as input of the MLP are then those columns of a given dictionary with higher absolute difference between frequency activation for each of the classes. That is, if some column is active many times for signals with AH events than for the signals without AH events, it is taken into account. Eq.3 shows the absolute difference of activation frequency F given two classes.

$$F = |\nu_{c1}^{j} - \nu_{c2}^{j}| \tag{3}$$

Finally MDAS method evaluate the performance of a MLP by varying its number of inputs and neurons in its hidden layer, next an optimal number of inputs P and neurons in hidden layer Q are selected.

2.3 INFERENCE ABOUT THE ACTIVATIONS COEFFICIENTS

Now a more general model is taken into account, where the observed signal vectors are noisy linear mixtures of M atoms of a dictionary. Thus, we have:

$$\mathbf{s} = \sum_{i=1}^{M} \phi_i a_i + \varepsilon_i = \Phi\mathbf{a} + \varepsilon. \tag{4}$$

Eq. 4 is associated to as a generative model, where a signal vector $\mathbf{s}$ is generated from a set of hidden sources a_j aligned as a representation vector $\mathbf{a}$, using a fixed dictionary Φ. The elements of ε are assumed to be uncorrelated, the analysis is then realized in terms of the noise covariance matrix $E\{\varepsilon^T\varepsilon\} = \Lambda_\varepsilon^{-1}$, with $E\{.\}$ denoting the expected value. The noise vector ε is then defined by $\varepsilon = \mathbf{s} - \Phi\mathbf{a}$ and the probability of s given the dictionary Φ and the hidden sources $\mathbf{a}$ can be stated as follows:

$$\pi(\mathbf{s}|\Phi, \mathbf{a}) = \exp(-\frac{\lambda}{2}|\mathbf{s} - \Phi\mathbf{a}|^2), \tag{5}$$

where $\lambda = 1/\sigma^2$ and σ is the standard deviation of the additive noise.

The elements of $\mathbf{a}$ are assumed to be statistically independent with a joint priori density:

$$\pi_{\text{prior}}(\mathbf{a}) = \prod_{j=1}^{M} \pi(a_j) \tag{6}$$

If Φ is known and $\mathbf{s}$ is given, the vector $\mathbf{a}$ can be estimated via the Bayes's formula by considering the posterior distribution (Eq. 7) and a maximum-a-posteriori (MAP) estimation of $\mathbf{a}$ is given by Eq. 8

$$\begin{aligned}\pi_{\text{post}}(\mathbf{a}|\Phi, \mathbf{s}) &= \frac{\pi(\mathbf{s}|\Phi, \mathbf{a})\pi_{\text{prior}}(\mathbf{a})}{\pi(\mathbf{s}|\Phi)} \\ &\propto \pi(\mathbf{s}|\Phi, \mathbf{a})\pi_{\text{prior}}(\mathbf{a})\end{aligned} \tag{7}$$

$$\begin{aligned} \mathbf{a}_{MAP} &= \underset{\mathbf{a}}{\operatorname{argmax}} \left[\pi_{\text{post}}(\mathbf{a}|\Phi, \mathbf{s}) \right] \\ &= \underset{\mathbf{a}}{\operatorname{argmax}} \left[\log \pi(\mathbf{s}|\Phi, \mathbf{a}) + \log \pi_{\text{prior}}(\mathbf{a}) \right] \end{aligned} \tag{8}$$

Lewicki and Olshausen [11] proposed an a priori distribution of Laplacian type, $\pi(a_j) = Ne^{\rho_j|a_j|}$, where $\rho_j < 0$ is given and this leads to the following rule for updating $\mathbf{a}$:

$$\Delta\mathbf{a} = \Phi^T \Lambda_\varepsilon \varepsilon - \rho^T |\mathbf{a}|, \tag{9}$$

where $\rho = (\rho_1, \rho_2, ..., \rho_n)^T$.

2.4 DICTIONARY LEARNING

In the previous subsection, a statistical method was used to solve the inference problem about activation coefficients. In what follows the learning problem is similarly solved. The atoms of Φ can be estimated by maximizing the log-likelihood function of the data given Φ [11], $L(\mathbf{s}, \Phi) \doteq E[\log \pi(\mathbf{s}|\Phi)]$, i.e. as follows:

$$\hat{\Phi} = \underset{\Phi}{\operatorname{argmax}} \left[L(\mathbf{s}, \Phi) \right]. \tag{10}$$

The log-likelihood function can be found by marginalizing the product of the conditional distribution of the data given the dictionary and the prior distribution of the coefficients. That is: $\pi(\mathbf{s}|\Phi) = \int_{\mathbb{R}^M} \pi(\mathbf{s}|\Phi, \mathbf{a}) \pi_{\text{prior}}(\mathbf{a}) \, d\mathbf{a}$.

The maximum in Eq. (10) can also be approximated by using a gradient ascent method with the following updating rule [12]:

$$\Delta\Phi = \eta \Lambda_\varepsilon E[(\mathbf{s} - \Phi\mathbf{a})\mathbf{a}^T]_{\pi_{\text{post}}(\mathbf{a}|\Phi, \mathbf{s})}, \tag{11}$$

where $\eta \in (0, 1)$ is the learning coefficient.

Although Eq. (11) provides an explicit updating rule for approximating the solution of (10), the RHS of (11) involves calculating the improper definite integral $\int_{\mathbb{R}^M} (\mathbf{s} - \Phi\mathbf{a})\mathbf{a}^T \pi_{\text{post}}(\mathbf{a}|\Phi, \mathbf{s}) \; d\mathbf{a} = E[(\mathbf{s} - \Phi\mathbf{a})\mathbf{a}^T]_{\pi_{\text{post}}(\mathbf{a}|\Phi, \mathbf{s})}$, which is almost always analytically impossible, and hence, it must be numerically approximated. As the dimension of $\mathbf{a}$ increases, however, this becomes a serious computational challenge and all traditional numerical methods turn out to be inadequate.

Lewicki and Sejnowski [13] used a multivariate Gaussian approximation to the posterior distribution around its maximum $\mathbf{a}_{MAP}$, $\pi_{\text{post}}(\mathbf{a}|\Phi, \mathbf{s}) \approx \sqrt{\frac{|H|}{2\pi^M}} e^{-1/2(\mathbf{a}-\mathbf{a}_{MAP})^T H (\mathbf{a}-\mathbf{a}_{MAP})}$, where H is the log-posterior evaluated at $\mathbf{a}_{MAP}$, $H = -\nabla\nabla^T \log \pi_{\text{post}}(\mathbf{a}|\Phi, \mathbf{s})_{MAP}$. Clearly the mean and covariance matrix of this Gaussian approximation are $\mathbf{a}_{MAP}$ and H^{-1}, respectively. With this approximation the updating rule (11) becomes:

$$\Delta\Phi = \eta \Lambda_\varepsilon ((\mathbf{s} - \Phi\mathbf{a})\mathbf{a}^T - \Phi H^{-1}). \tag{12}$$

In order to obtain Φ and the coefficients (Eqs. (9) and (12)), the implementation proposed by Lewicki and Olshausen [11] was used at the dictionary training stage.

2.5 INVERSE PROBLEMS

The problem for finding the activation coefficients $\mathbf{a}$, given Φ and $\mathbf{s}$, gives rise to an Inverse Problem (IP) with sparse constraints. A greedy algorithm called Orthogonal Matching Pursuit (OMP) is chosen in order to find a solution of IP. The reason for choosing this algorithm is because it provides a good sparse approximated solution faster than most other methods to solve (2).

The OMP algorithm proposes a modification of the greedy Matching Pursuit algorithm (MP) of Mallat and Zhang [14], where the full backward orthogonality of the residual convergence is maintained. It is shown that all additional computation required for the OMP algorithm may be performed recursively [15].

Mallat and Zhang show a sequence of approximations used for the MP algorithm in order to obtain an SR of the signal $\mathbf{s}$. It is assumed that $\mathbf{a}$ has only m non-zero components, and therefore the signal vector $\mathbf{s} = \Phi\mathbf{a}$ is a linear combination of m columns from Φ. To identify one such a vector $\mathbf{a}$, it is necessary to distinguish which columns of Φ participate in the measured signal $\mathbf{s}$. The idea of the OMP algorithm basically consists in properly selecting columns of Φ. At each iteration, the column most correlated with the current residual of $\mathbf{s}$ is taken. Then the residual is updated and iterated. In this way, after m iterations the algorithm will choose a set of m columns of Φ [16].

3 EXPERIMENTS

The Sleep Heart Health Study (SHHS) database [1] is used for this work. This database contains exhaustive information about detailed studies which are appropriately designed to investigate the relationships between sleep breathing disorders and cardiovascular diseases. The full dataset contains nearly 1000 complete PSGs, each one of them containing several biomedical signals such as ECG, nasal airflow, respiratory effort and SaO_2, among others. Annotations of sleep stages, arousals and respiratory events (apnea and hypopnea) are also included. Only the SaO_2 signals and the AH events will be of our interest.

First of all, a wavelet processing technique is used for denoising the SaO_2 signal, which was sampled at 1Hz. The denoised SaO_2 signal is then obtained by making zero the approximation coefficients, at level 8, of the discrete Dyadic Wavelet Transform (DWT) with a mother function Daubechies 2 [17]. In the sequel, the SaO_2 signal shall always refer to the denoised one.

For this work, a subset of 84 studies are selected in order to analyze the performance of a screening method for sleep disorders [10]. The dataset is divided into training and test sets consisting of 20 and 64 studies, respectively. The training set contains 4 groups of 5 PSGs corresponding to AHI values below 5, between 5 and 10, between 10 and 15 and above 15. The test set comprises 64 PSGs with different degrees of illness. For both training and test sets, each SaO_2 signal is segmented into vectors of length 128, with an overlapping of 32 elements. The segments are then arranged as column vectors $\mathbf{s}_j \in \mathbb{R}^{128}$.

[1]Database: http://physionet.org/physiobank/database/shhpsgdb/

These segments are also labeled as belonging to *class 1* or *class 2* depending on whether they contain AH events or not, respectively.

Next, we construct two class-training matrices S^{c1}_{train} and S^{c2}_{train} by stacking side-by-side all vectors $\mathbf{s}$ labeled as *class 1* and *class 2* in the training set, respectively. The matrix S_{train} is then defined as $S_{train} = [S^{c1}_{train}\ S^{c2}_{train}]$, while the matrix S_{test} is built by stacking side-by-side all vectors $\mathbf{s}$ in the test set.

In the next step two complete dictionaries Φ_{c1} and Φ_{c2} are learned by using matrices S^{c1}_{train} and S^{c2}_{train}, respectively. An overcomplete dictionary Φ_1 is then built as $\Phi_1 = [\Phi_{c1}\ \Phi_{c2}]$. Also, a complete dictionary Φ_2 is learned without taking into account class information, i.e. by using the whole matrix S_{train}. In all cases the learning process is performed by means of the Noise Overcomplete ICA (NOCICA) [13] method.

Now given $\mathbf{s} \in \mathbb{R}^M$, a constant $k \in \mathbb{N}$ and a dictionary Φ, we formulate the following linear inverse problem with sparse constraint:

$$\begin{aligned} J_{\Phi,\mathbf{s}}(\mathbf{v}) &\doteq ||\mathbf{s} - \Phi\mathbf{v}||_2^2 \\ \mathbf{v}(\Phi, \mathbf{s}, k) &\doteq \underset{\mathbf{v}\in\mathbb{R}^M, ||\mathbf{a}||_0 \le k}{\operatorname{argmin}} \ J_{\Phi,\mathbf{s}}(\mathbf{v}). \end{aligned} \tag{13}$$

We denote by $\mathbf{c_1} \doteq \mathbf{v}(\Phi_1, \mathbf{s}_j, k)$ and $\mathbf{c_2} \doteq \mathbf{v}(\Phi_2, \mathbf{s}_j, k)$ the solutions of (13) for $\mathbf{s} = \mathbf{s}_j$, where $\mathbf{s}_j$ is the j^{th} column of S_{train}. In the same way, we denote by $\mathbf{d_1} \doteq \mathbf{v}(\Phi_1, \mathbf{s}_j, k)$ and $\mathbf{d_2} \doteq \mathbf{v}(\Phi_2, \mathbf{s}_j, k)$ the solutions of (13) for $\mathbf{s} = \mathbf{s}_j$, where $\mathbf{s}_j$ is the j^{th} column of S_{test}.

Finally, for the classification step, the MDAS method described in Section 2.2 is then used in order to select the optimal number of the MLP inputs and neurons in its hidden layer for both dictionaries Φ_1 and Φ_2.

Two classification performance measures are utilized to compare the use of both complete (Φ_2) and overcomplete (Φ_1) dictionaries. For AH events detection, the *sensitivity* (SE) is defined as the proportion of segments with AH events for whom an event is present, while *specificity* (SP) is defined as the proportion of segments without AH events for whom an event is not present. Now for moderate or severe OSAHS diagnoses, the *sensitivity* is defined as the proportion of patients with disease for whom the test is positive, while *specificity* is defined as the proportion of patients without disease for whom the test is negative. Also, a Receiver Operating Characteristics (ROC) [18] analysis is made, which the following parameters are obtained. True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN), cut-off point and Area Under the Curve (AUC).

The detection of patients with a moderate or severe OSAHS is clinically very important. For this reason, in this work an AHI of 15 or more shall be considered as having a positive diagnostic for OSAHS.

4 RESULTS

In the results presented below we used a sparsity level $k = 16$. Also, for improving the performance of the classifiers, we took the same number of segments in each one of both classes. For all 4 methods the optimization problem (13) was solved via OMP algorithm due to its high efficiency and effectiveness. Note that the OMP algorithm has to go through 16 iterations to calculate the solution of (13) with $||\mathbf{a}||_0 = 16$.

In a previous work [10], the performance of an MLP was analyzed by varying its number of inputs and its number of neurons in hidden layer from 2 to 100 and from 10 to 20, respectively. Now, both full vectors $\mathbf{c_1}$ and $\mathbf{c_2}$ are taken into account for training the MLP (Table 1). The MLP outputs were labeled as "1" or "0" depending on whether an AH event was detected or not, respectively. Finally the estimated AHI (AHIest) was obtained by the total number of "1" divided by the time duration of each study (in hours).

Table 1: Number of inputs and neurons in hidden layer used for training the MLP.

Method	# inputs	# neurons in hidden layer
MDAS-OAD	24	14
MDAS-CD	30	14
OAD	256	14
CD	128	14

Next the whole vectors $\mathbf{d_1}$ and $\mathbf{d_2}$ were taking into account in order to analyze the performance of the MLP for both overcomplete and complete cases, respectively. Table 2 shows the sensitivity and specificity measures taking account the AH events classifications as well as the corresponding AHIest-AHI correlation percentages. As seen in Table 2, significantly high correlation percentages for detecting AH events were obtained. Among the 64 studies contained in the test set, 48 patients without presence of moderate or severe OSAHS were diagnosed by the expert while 16 studies for whom moderate or severe OSAHS were diagnosed by the expert.

Table 2: Classifier's performance for AH events detection

Method	SE	SP	Correlation percentage
MDAS-OAD	74,52%	76,73%	90,04%
MDAS-CD	68,86%	67,69%	74,57%
OAD	62,84%	65,63%	91,24%
CD	61,13%	64,85%	92,23%

The ROC curves for the diagnostic accuracy for moderate or severe OSAHS detection when using MDAS method are shown in Fig. 1, while Fig. 2 shows the diagnostic accuracy for moderate or severe OSAHS without applying MDAS method.

Table 3 shows the values of the performance measures for OSAHS detection obtained for all 4 methods. First with the MDAS-OAD method, a total of 15 patients were correctly detected as having OSAHS and 44 patients were also correctly diagnosed without

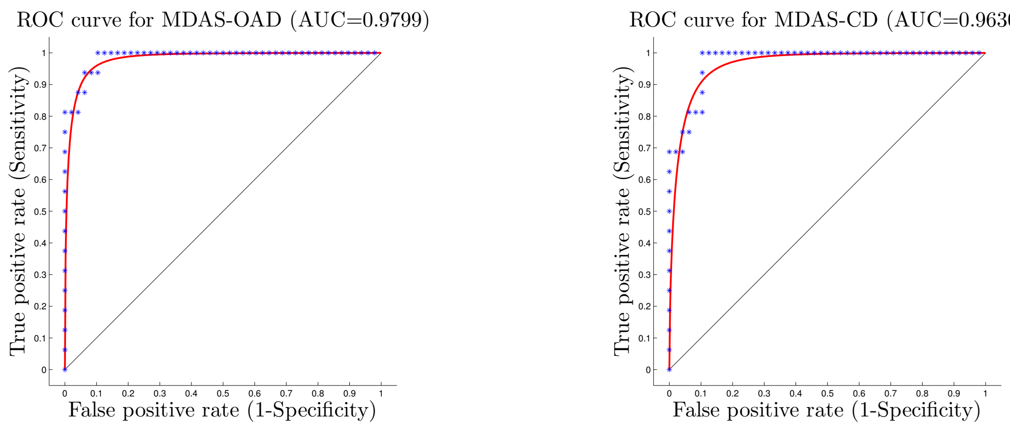

Figure 1: ROC curves for detecting moderate or severe OSAHS, MDAS-OAD (left) and MDAS-CD (right).

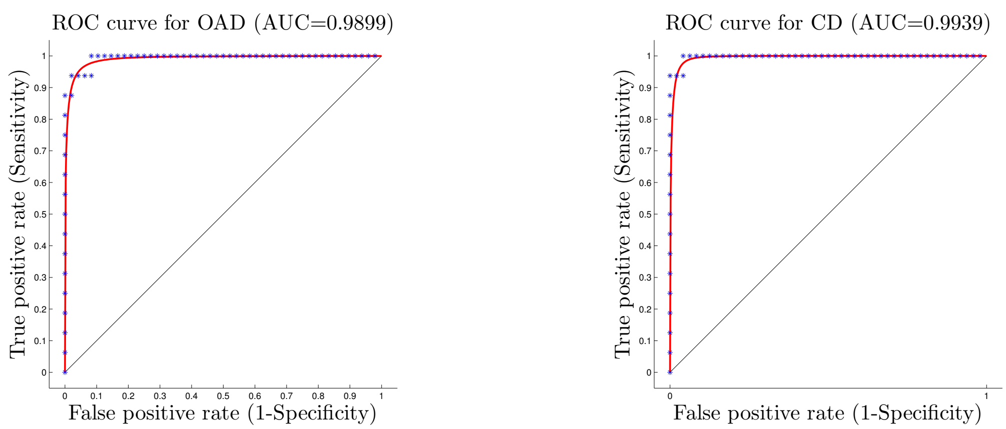

Figure 2: ROC curves for detecting moderate or severe OSAHS, OAD (left) and CD (right).

presence of OSAHS. Moreover this method resulted in 4 false positive and 1 false negative diagnoses. With the MDAS-CD method, all 16 patients with OSAHS were detected and 43 patients were correctly diagnosed without presence of OSAHS. This method resulted in 5 false positives and no false negative diagnoses. Analogously, the performance measures obtained with the OAD and CD methods are shown in the last two rows of Table 3.

Locking at the last column the Table 3 we see the AUC values obtained with all 4 methods are significantly high, corresponding the highest value to the CD method (AUC=0,9939). Table 3 also shows that although the application of the MDAS method has a negative impact in the specificity measure, it has no effect on the sensibility. Similarly the specificity for MDAS-OAD and OAD methods resulted higher than specificity obtained with MDAS-CD and CD methods, respectively.

Table 3: Performance measures for diagnosing moderate or severe OSAHS.

Method	TP	TN	FP	FN	SE	SP	Cut-off	AUC
MDAS-OAD	15	44	4	1	93,75%	91,67%	399,65	0,9799
MDAS-CD	16	43	5	0	100%	89,58%	5.8382	0,8958
OAD	15	46	2	1	93,75%	97,92%	306,85	0,9899
CD	16	46	2	0	100%	95,83%	6.66	0,9939

5 CONCLUSIONS

The previous analysis shows that the sparse representation of an SaO_2 signal is a suitable technique for detecting moderate or severe OSAHS. The algorithm NOCICA was found to be a very useful tool for learning both overcomplete and complete dictionaries. Also, to linearly regress the apnea-hypopnea events a multilayer perceptron was successfully constructed.

The obtained results show that although the sensibility and specificity values (Table 2) were not so good, a high correlation between the AHI observed by the expert via PSG and the AHIest obtained by using the sparse representation of the SaO_2 signal. This fact constitutes a strong evidence that such a procedure could be successfully used for detecting moderate or severe OSAHS. Also, as shows Table 3 a very high performance for moderate or severe OSAHS diagnosing was achieved.

Although similar classification results were obtained with and without applying the MDAS method, it is important to point out that the use of this procedure reduces highly the dimensionality of the problem, and therefore the CPU time required for classification.

As mentioned in Section 4, a sparsity level of $k = 16$ was selected. For future works we propose to vary k from a low to a high level and analyze its impact on the diagnostic performance measures. Also, the number of inputs P and the threshold of the MLP output will be further studied.

6 ACKNOWLEDGEMENTS

This work was supported in part by Consejo Nacional de Investigaciones Científicas y Técnicas, CONICET, through PIP 2014-2016 Nro. 11220130100216-CO and PIP 2012-2014 Nro. 114 20110100284-KA4, by the Air Force Office of Scientific Research, AFOSR / SOARD, through Grant FA9550-14-1-0130 and by Universidad Nacional del Litoral through project CAI+D 501 201101 00519 LI.

REFERENCES

[1] Thorpy M.J. *The international classification of sleep disorders, Revised. Diagnostic and coding manual.* American Academy of Sleep Medicine, (2001).

[2] Nogueira F., Nigro C., Cambursano H., Borsini E., Silio J. and Ávila J. *Guías prácticas de diagnóstico y tratamiento del síndrome de apneas e hipopneas obstructivas del sueño*. Medicina, 349-362, Vol. 73, (2013).

[3] Durán-Cantolla J. *et al Consenso Nacional sobre el síndrome de apneas-hipopneas del sueño*, 1-110, Vol. 41, (2005).

[4] Larrateguy L.D. *Síndrome de Apneas-Hipopneas Obstructivas del Sueño (SAHOS). Manifestaciones clínicas. Diagnóstico*. Medicina del Sueño. Editorial Mediterráneo. Santiago. Chile., (2007).

[5] Nieto F.K., Young T.B. and Lind B.K. *Association of sleep-disordered breathing, sleep apnea, and hypertension in a large community-based study*. JAMA, 1829-1836, (2000).

[6] Newman A.B., Nieto J. and Guirdry U. *Relation of sleep-disordered breathing to cardiovascular risk factors*. The Sleep Hearth Healt Study. Am J Epidemiol, Vol. 154, 50-59, (2001).

[7] Hyvrinen A. *Survey on independent component analysis*. Neural Computing Surveys, 94-128, Vol. 2, (1999).

[8] Olshausen B.A. and Field, D.J. *Vision and the coding of natural images*. American Scientist, 238-245, Vol. 88, (2000).

[9] Lestussi F.E., Rufiner H.L., Di Persia L.E. and Milone D.H. *Sparse coding for apnea-hypopnea detection*. XIV Reunión de Trabajo en Procesamiento de la Información y Control (RPIC), 1-6, (2011).

[10] Rolon R.E. (et al) *Most Discriminative Atom Selection for apnea-hypopnea events detection*. Anales del VI Congreso Latinoamericano de Ingeniería Biomédica, 709:712, (2014).

[11] Lewicki M.S. and Sejnowski T.J. *A probabilistic framework for the adaptation and comparison of image codes*. Optical Society of America, 1587-1601, Vol. 16, (1999).

[12] Abdallah, S.A. *Towards music perception by redundancy reduction and unsupervised learning in probabilistic models*. Ph.D. Thesis, Department of Electronic Engineering, King's College London, (2002).

[13] Lewicki M.S. and Sejnowski T.J. *Learning overcomplete representations*. Advances in Neural Information Processing 10 (Proceedings NIPS97), 556-562, (1998).

[14] Mallat S.G. and Zhang Z. *Matching pursuits with time-frequency dictionaries*. IEEE Transactions on Signal Processing, Vol. 41, 3397-3415, (1993).

[15] Pati Y.C., Rezaiifar R. and Krishnaprasad P.S. *Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition*. IEEE Asilomar Conference on Signals, Systems and Computers, Vol. 1, 40-44, (1993).

[16] Tropp J.A. and Gilbert A.C. *Signal recovery from random measurements via orthogonal matching pursuit*. IEEE Transactions on Information Theory, Vol. 53, 4655-4666, (2007).

[17] Lestussi F.E., Di Persia L.E. and Milone D.H. *Comparison of on-line wavelet analysis and reconstruction: with application to ECG*. 5th International Conference on Bioinformatics and Biomedical Engineering, (iCBBE), 1:4, (2011).

[18] Kumar, R. and Indrayan, A. *Receiver operating characteristic (ROC) curve for medical researchers*. Indian Pediatrics, Vol. 48, 277-287, (2011).